ED 447 147   TM 031 856

AUTHOR	Russell, Michael
TITLE	Using Expected Growth Size Estimates To Summarize Test Score Changes. ERIC/AE Digest.
INSTITUTION	ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	EDO-TM-00-04
PUB DATE	2000-09-00
NOTE	4p.; Based on "Using Expected Growth Size Estimates To Summarize Test Score Changes" by Michael Russell, "Practical Assessment, Research & Evaluation," 7(6).
CONTRACT	ED99CO0032
AVAILABLE FROM	For full text: http://ericae.net/pare/getvn.asp?v=7&n=6.
PUB TYPE	ERIC Publications (071) -- ERIC Digests in Full Text (073)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Achievement Gains; Change; Effect Size; *Evaluation Methods; *Scores; Test Interpretation; *Test Results
IDENTIFIERS	ERIC Digests; *Expected Growth

ABSTRACT

An earlier Digest described the shortcomings of three methods commonly used to summarize changes in test scores. This Digest describes two less commonly used approaches for examining changes in test scores, those of Standardized Growth Estimates and Effect Sizes. Aspects of these two approaches are combined and applied to the Iowa Test of Basic Skills to demonstrate the usefulness of a third method, termed Expected Growth Size, to examine change in test scores. An expected growth size is more difficult to calculate than the other methods, but it offers three advantages. By expressing change in relation to the standard deviation, growth rates for different tests and different grade levels can be compared directly. Once expected growth sizes are calculated for a given test, they can be transformed easily to more common measurement scales. And once expected growth sizes are transformed to a Normal Curve Equivalent scale, changes in an individual's or a group's mean score can be reported in relation to expected growth. How to calculate expected growth size is illustrated. (SLD)

Using Expected Growth Size Estimates to Summarize Test Score Changes

Michael Russell
Boston College

Using Expected Growth Size Estimates to Summarize Test Score Changes

Michael Russell, Boston College

An earlier digest described the shortcomings of three methods commonly used to summarize changes in test scores (Russell, 2000). This article describes two less commonly used approaches for examining change in test scores, namely Standardized Growth Estimates and Effect Sizes. Aspects of these two approaches are combined and applied to the Iowa Test of Basic Skills (ITBS) to demonstrate the utility of using a third method, termed Expected Growth Size, to examine change in test scores.

Standardized Growth Estimates

Stenner, Hunter, Bland, & Cooper describe a standardized growth expectation (SGE) as "the amount of growth (expressed in standard deviation form) that a student must demonstrate over a given treatment interval to maintain his/her relative standing in the norm group" (1978, p. 1). To determine an SGE, Stenner et. al. proposed the following three-step method.

Step 1. The scale score associated with the 50th percentile for a given grade level or the pre-test is identified.

Step 2. The percentile rank for the following grade level or the post-test associated with this scale score is found.

Step 3. The difference between the 50th percentile and the post-test percentile is calculated.

To determine this difference, a unit normal deviate table is used to convert percentiles to z-scores and the z-score for the post-test is subtracted from the z-score for the pre-test.

The difference between the pre-test and post-test z-scores is the SGE and expresses "the amount of loss in relative standing that such a student would suffer if he/she learned nothing during the time period" (Stenner, et. al., 1977, p. 1).

As an example, to determine the SGE for grade 3, Table 1 indicates that the scale score associated with the 50th percentile for grade 3 on the ITBS Language sub-test is 174. The percentile rank for grade 4 that corresponds to a scale score of 174 is 26. If a student received the same scale score in grades 3 and 4, their percentile rank would drop from 50 to 26. After both percentiles are converted to z-scores and subtracted, the difference between the two z-scores represents the SGE. In this case, the z-scores corresponding to percentile ranks of 50 and 26 are 0 and -.64, respectively. Thus, the SGE is .64, which indicates a relative loss of .64 standard deviations for a student who shows no change in his/her test score.

Effect Sizes

When applying Stenner et. al.'s method for calculating SGEs, Haney, Madaus and Lyons (1993, p. 231-32) point out that the idea of a SGE is analogous to an effect size in that each represents the difference in mean performance of two groups expressed in standard scores. As Glass, McGaw and Smith (1981) describe, an effect size represents the difference between two groups in standard deviations. To calculate an effect size, the difference between the mean of the control group and the experimental group is divided by the standard deviation of the control group. Conceptually, the only difference between an effect size and an SGE is that an effect size is used to compare the means of a "control" group and an "experimental" group while a SGE compares the performance of groups of students at various grade levels.

Table 1: Percentile Rank, Standard Score and Standard Deviations for the Iowa Test of Basic Skills Language Sub-test

Standard Score	Percentile Rank	
	Grade 3	Grade 4
174	50	26
175	52	27
176	54	29
...	...	...
189	78	47
190	79	48
191	81	50
St. Dev.	19.05	24.25

In the SGE example above, the third grade is designated as the control group and the fourth grade is the experimental group. To determine the effect size or amount of growth between grade three and grade four, the standard score associated with the 50th percentile rank for grade three is subtracted from the standard score associated with the same percentile rank for grade four. This difference is divided by the standard deviation for grade three. Focusing on Table 1, the effect size for grade three is found by subtracting 174 from 191 and dividing by 19.05. The resulting effect size indicates that a student's test score must increase by .89 standard deviations to maintain his/her standing at the 50th percentile.

Expected Growth Size

Although an SGE and an effect size are similar, there is one important difference: an SGE focuses on the standing lost when there is no change in test score, while the effect size focuses on the amount of change in a test score necessary to maintain one's standing. When applied in this manner, the effect size method provides an estimate of the expected growth size between two time periods. In the example above, the expected growth size (EGS) between grade three and grade four on the ITBS Composite Language test is .89 standard deviations.

Defining the Base Year or Control Group

In a well-designed experiment, there is little question as to which group is defined as the control group and which is the experimental group. However, when applying the concept of an effect size to change in test scores between two grade levels, one could reference growth to the pre-test or the post-test distribution.

In the case of SGEs, the post-test distribution is used to reference "growth". Note, however, that although SGEs employ the term growth, the methodology actual provides a measure of loss assuming that a student experiences no growth whatsoever. In this way, using the post-test distribution to reference "growth" is fundamentally flawed in that change is placed in the context of where a student is expected to be rather than from where they started. The situation is analogous to describing someone's progress on trip in relation to how far they still must go in order to reach their destination rather than from how far they have traveled since their departure.

In the case of using an effect size to express growth between two grade levels, one might argue that the pooled standard deviation be employed in lieu of the standard deviation of the control group. However, the difficulty of obtaining an estimate of the pooled standard deviation for most standardized tests forces a choice between designating the pre-test or the post-test as the control group. Given the desire to measure change or growth from where a group begins at one

point in time to where they end at a second point in time, the EGS methodology references change to the pre-test distribution. For this reason, the pre-test distribution is assigned as the control group.

Advantages of an Expected Growth Size

Although an expected growth size is more difficult to calculate, it offers three advantages. First, by expressing change in relation to the standard deviation, growth rates for different tests and different grade levels can be compared directly. Table 2 presents expected growth sizes for grades 1 through 8 for several portions of the ITBS. Examining Table 2, one can see that the expected growth sizes differ for each portion of the ITBS. Table 2 also shows an inverse relationship between grade level and size of expected growth. As the grade level increases, the amount of growth students experience decreases.

Table 2: Expected Growth Sizes for the ITBS Reading, Language, Math and Composite Tests

Grade Level	Growth Size for the 50th Percentile			
	Reading	Language	Math	Composite
1	NA	1.46	1.38	1.69
2	.93	1.10	1.25	1.24
3	.79	.89	.89	.99
4	.67	.58	.68	.73
5	.52	.50	.53	.54
6	.39	.32	.42	.43
7	.39	.29	.38	.36
8	.36	.29	.40	.40

Similarly, within each grade level, the amount of growth students experience varies by percentile ranks. Students scoring at the 25th percentile experience less growth than students scoring at the mean. And students scoring at the mean experience less growth than students scoring at the 75th percentile. This pattern explains why the standard deviation for most standardized tests increases as the grade level progresses.

Second, once expected growth sizes are calculated for a given test, they can be easily transformed to more common measurement scales. As an example, multiplying the expected growth size by the standard deviation of an Normal Curve Equivalent, NCE, (21.06) provides the number of NCE points a student's score increases during a given time period relative to the student's initial norm group when s/he maintains his/her current standing. For the ITBS Language test, the score for a student who maintains a 50th percentile ranking increases 18.74 NCEs between the third and fourth grade.

Third, once expected growth sizes are transformed to an NCE scale, changes in an individual's or a group's mean score can be reported in relation to expected growth. Performance on most standardized tests is reported relative to the Norm Group for a student's current grade. If the student grows at the same rate as other students in the Norm Group, his/her percentile rank and NCE will remain the same across two years. However, if the student's rate of growth differs from that of the Norm Group, his/her NCE and percentile rank will change.

The expected growth size can be used to determine the extent to which the student's growth exceeded or fell short of the expected growth size. To do so, the student's current NCE is subtracted from his/her previous NCE and divided by the expected NCE growth rate. As an example, consider a student whose NCE for the ITBS Language test increased from 50 in grade 3 to 55 in grade 4. When divided by the expected NCE growth size for third grade (18.74), this five point increase represents 1.27 years of growth. Thus, the student's score increased 27% more than expected.

As Table 2 indicates, growth sizes vary across grade levels. Expressing change in test scores in relation to expected growth size

takes these differences in growth rates into consideration. The extent to which performance changes is placed in the context of how scores generally change for students in a given grade. As a result, a more accurate measure of how a student changes relative to other students in his/her grade is produced. As an example, Table 2 shows that students in grade 2 experience about twice as much growth in their test scores compared to students in grade 5. For this reason, an increase of 5 NCEs on the ITBS Composite Math test represents larger growth relative to expected growth for a student in grade 5 than for a student in grade 2.

Limitations of Expected Growth Sizes

Although expected growth sizes provide a sounder approach for summarizing change in test scores than some of the more commonly used approaches, their use is limited to norm referenced standardized tests. Moreover, the EGS methodology assumes that the tests have been vertically equated. When comparing change across multiple years, the methodology also assumes that the tests administered each year provide measures of the same construct based on identical content. Although most norm-referenced tests attempt to meet both assumptions – vertical equating and measures of the same construct – the extent to which they fail to meet these assumptions impacts the accuracy of estimates yielded by the EGS methodology. Finally, as with all comparisons of change over time, the EGS method is also limited by the reliability of the scores used to calculate change. Although there is considerable debate over the extent to which low score reliability impacts the meaningfulness of change scores, caution is advised when employing the EGS method for tests with low reliability (see Willet, 1988 for fuller discussion on reliability and change scores).

Using Expected Growth Sizes for Your Students

To apply expected growth sizes to examine change in the performance of your students, readers are encouraged to use the attached spreadsheet. The spreadsheet provides an easy-to-use template that allows users to calculate expected growth sizes for most standardized tests. In addition, the spreadsheet translates expected growth sizes into expected changes in NCE scores for each grade level.

As the attached instructions indicate, two pieces of information are required to use the spreadsheet: 1. Standard Score to Percentile Rank Conversion tables for the standardized test; and 2. The standard deviation for the standard score for each grade level. This information is available in the Technical Report(s) for each standardized test.

Although expected growth sizes are more complicated to calculate, they provide a more accurate and comparable method of examining change in test scores within and across grade levels and on different tests.

References

Glass, G., McGaw, B. & Smith, M. L. (1981). *Meta-analysis in Social Research.* Beverly Hills: Sage.

Haney, W., Madaus, G., & Lyons, R. (1993). *The Fractured Marketplace for Standardized Testing.* Boston, MA: Kluwer Academic Publishers

Russell, M. (2000). Summarizing change in test scores: shortcomings of three common methods. *ERIC Digest Series.* Also in *Practical Assessment, Research and Evaluation,* 7(5). [Available online: http://ericae.net/pare/getvn.asp?v=7&n=5].

Stenner, A. J., Hunter, E. L., Bland, J. D., & Cooper, M. L. (1978). *The standardized growth expectation: Implications for educational evaluation.* Paper presented at the Annual Conference of the American Educational Research Association, Toronto, Canada. (ERIC Document Reproduction Service Number ED169072.)

Willett, J. (1988). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of Research in Education 15* (pp. 345-422). Washington, DC: American Educational Research Association.

4